DQN拡張手法の有効性検証

～Atari SpaceInvadersにおけるAblation Study～

深層強化学習講座2025Summer　最終課題

yf591

2025年9月29日

背景と目的

【背景】
- 深層強化学習の基礎アルゴリズムDQNには、「Q値の過大評価」により学習が不安定になる課題がある。
- その解決策としてDouble DQNやPrioritized Experience Replay (PER)などの拡張手法が提案されている。

【目的】
- これらの拡張手法が、学習の「何」を「どれくらい」改善するのかを定量的に明らかにすること。

【手法】
- ベースラインのDQNに対し、拡張手法を一つずつ追加していくAblation Study（除去実験）を実施

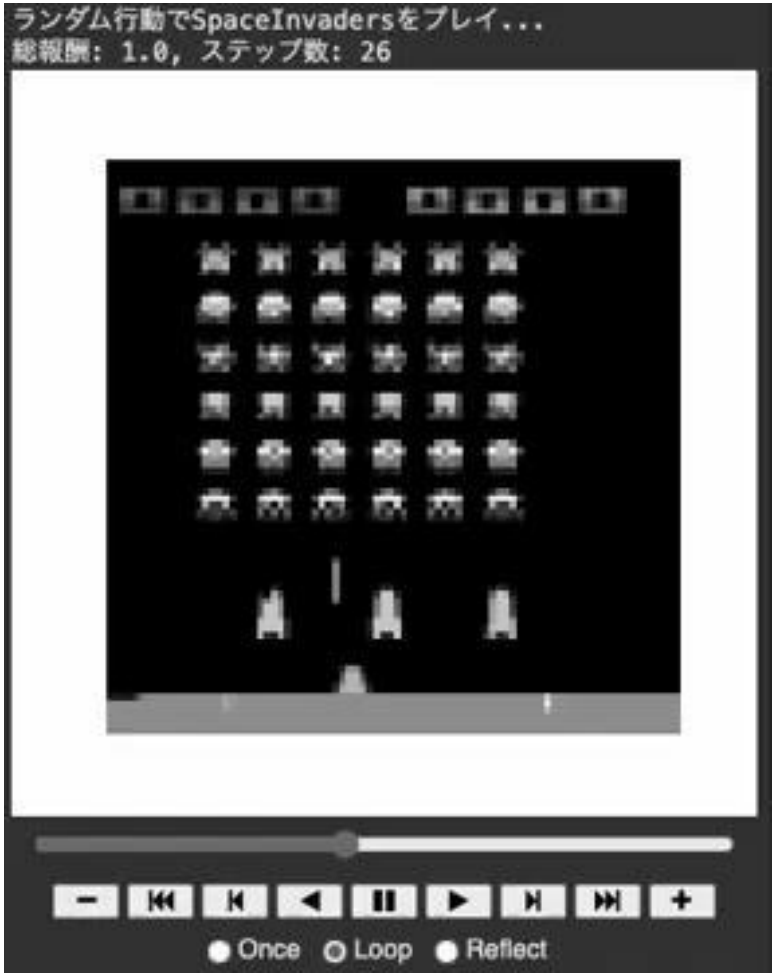
実験の設定

【環境】
Atari SpaceInvaders-v5（右の動画）

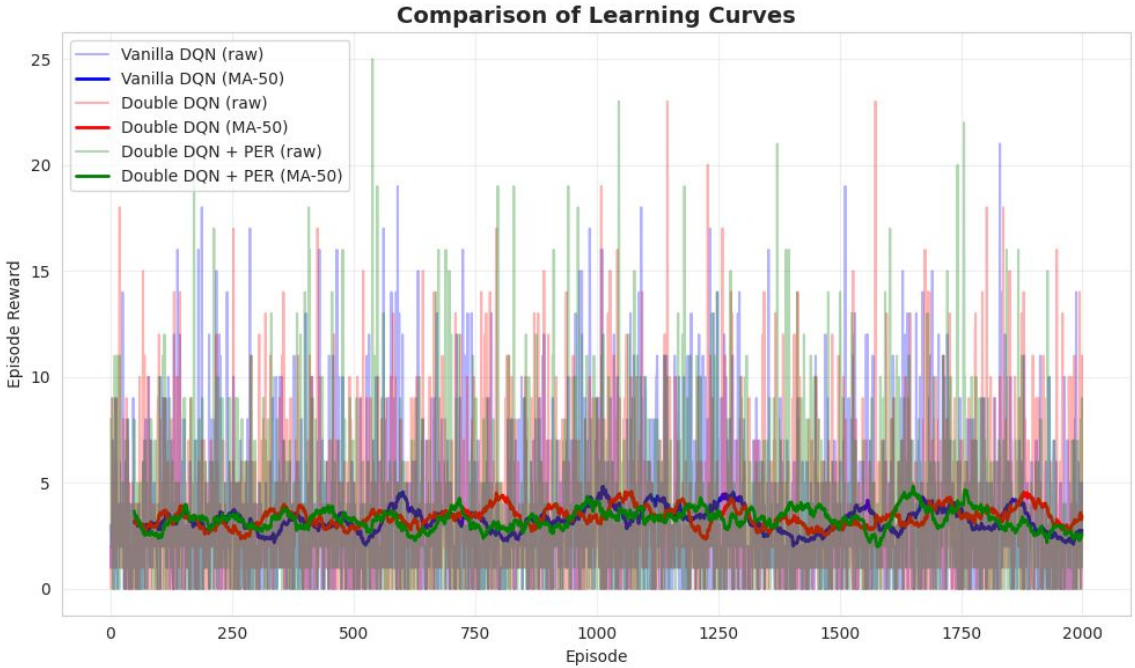
【比較モデル】
1. Vanilla DQN (ベースライン)
2. DQN + Double DQN
3. DQN + Double DQN + PER

【評価指標】
- 最終性能（学習終盤の平均スコア）
- 学習速度（目標スコアへの到達時間）
- 安定性（学習中のスコアのばらつき）

【学習ステップ】
- 各モデル 最大100万タイムステップが理想（実験では2000エピソード、約8万タイムステップ）



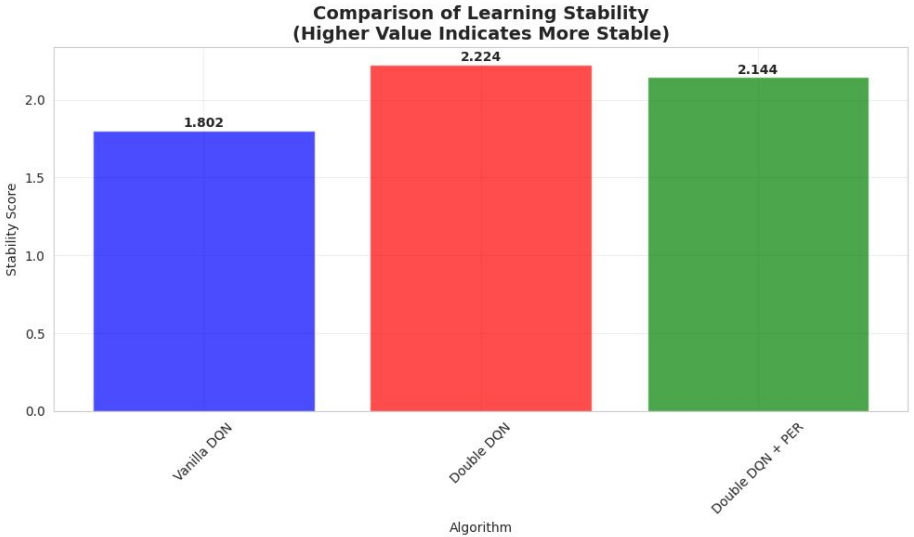
結果①　学習速度と安定性の改善について

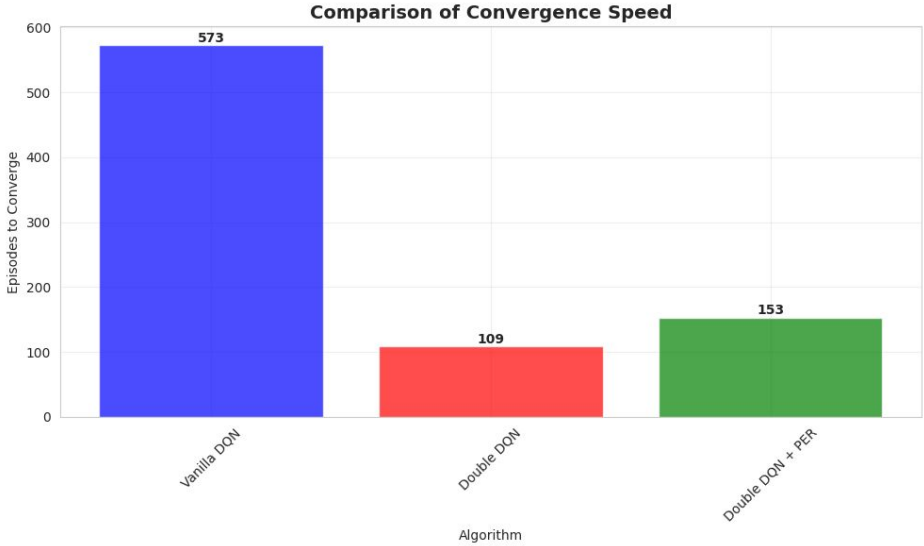


【結果】

- 学習曲線では、3手法とも移動平均線が明確な右肩上がりを描くまでには至らなかった。
- 実験で実施した2000エピソードの段階ではまだ性能が飽和するには程遠い状態であると考えられる。

結果①　学習速度と安定性の改善について





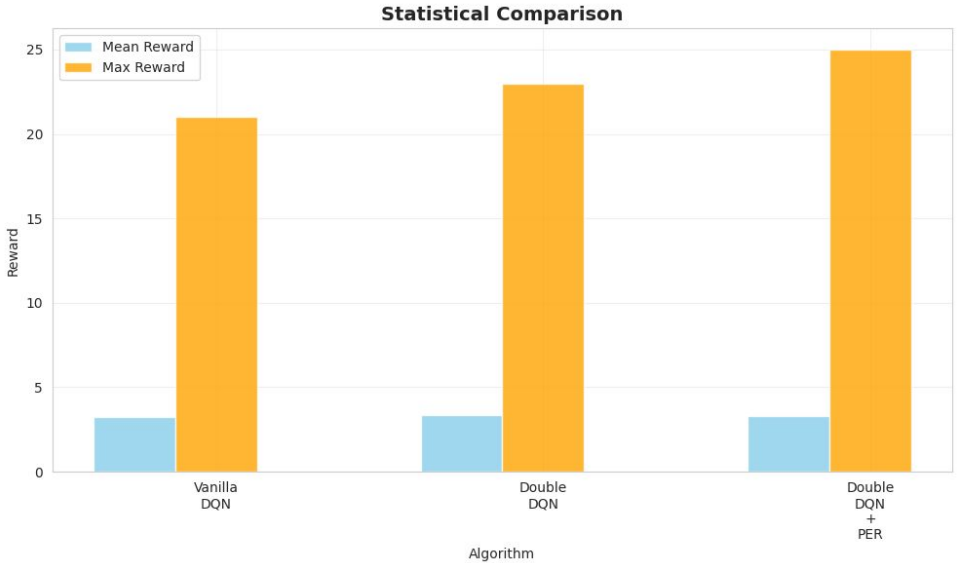
【結果】

- **学習安定性**は**Double DQNの安定性スコアが 2.224と最も高く**、学習全体の安定化につながっていると確認できた。
- Double DQNの導入によって、**学習速度が 5倍以上向上**した。
  Vanilla DQN **573 episodes** ▶▶▶ Double DQN **109 episodes**

結果②　最終性能とポテンシャルの比較

性能比較テーブル

アルゴリズム	直近100平均	収束エピソード	最大報酬
Vanilla DQN	2.6	573	21
Double DQN	**3.3**	**109**	23
Double DQN + PER	2.8	153	**25**

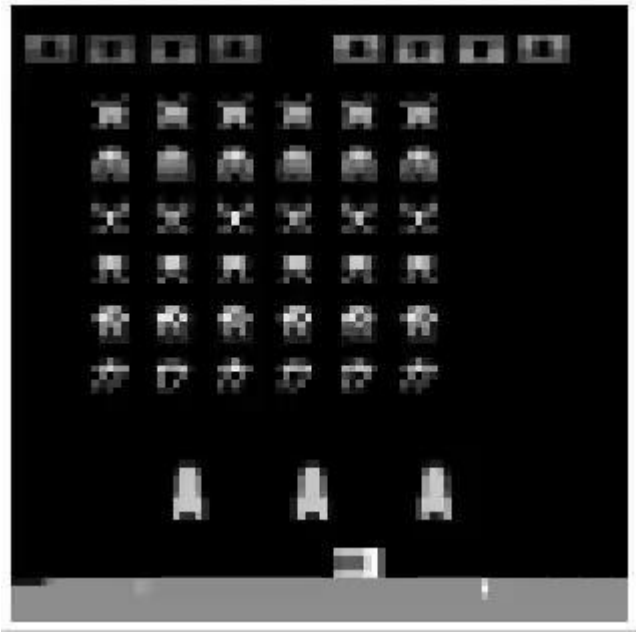


【結果】

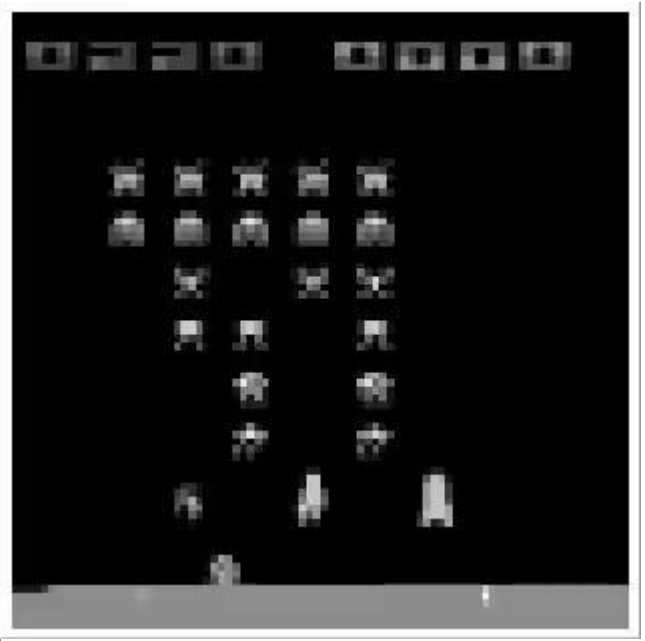
- **安定した最終性能は Double DQN**が最高
  学習終盤の平均スコア（Last 100 Mean）が3.3と最も高かった
- **瞬間最大スコアは DQN+PER**が記録
  Max Rewardが25を達成し、最高のポテンシャルを示した

【参考】プレイ動画比較

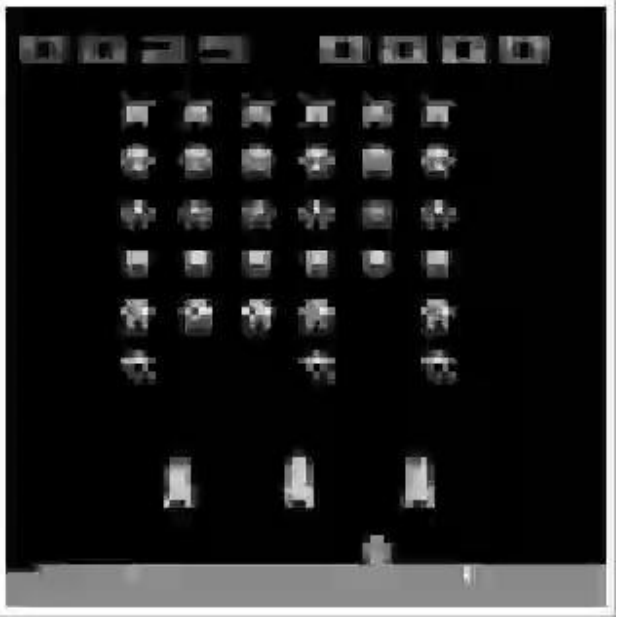
ベースモデル



Double DQN



Double DQN + PER



まとめ & 今後の展望（課題）

【本実験のまとめ】

1. AtariのSpaceInvaders（スペースインベーダー）環境において、Double DQNは、学習の速度・安定性・最終性能の全てを改善する極めて有効な手法であることを確認できた。

2. PERの追加は、学習初期では安定性を損なう一方、瞬間的なハイスコアを出すポテンシャルを示した。

【今後の展望（課題）】

- 長期学習の実施

  今回の実験では約8万ステップ（max_episodes 2000）までしか試せなかったが、100万ステップまで学習を継続し、PERの真価を検証する。

- 他の拡張手法の検証

  Dueling Networkなど、他の手法との組み合わせ効果も検証したい。